

Towards Retrieving Force Feedback in Robotic-Assisted Surgery: A Supervised Neuro-Recurrent-Vision Approach

Angelica I. Aviles, *Student Member, IEEE*, Samar Alsaleh, *Student Member, IEEE*,
James K. Hahn, and Alicia Casals, *Senior, IEEE*

Abstract—Robotic-assisted minimally invasive surgeries have gained a lot of popularity over conventional procedures due to many benefits to both surgeons and patients. Nonetheless, they still suffer from one major limitation: the lack of force feedback which restricts the surgeon's sense of touch and may lead to tissue damage during a procedure. To overcome this limitation, we propose a novel force estimation approach that combines visual information with supervised learning to estimate the applied force and provide the surgeon with a suitable representation of the tissue deformation. The proposed solution starts with extracting the geometry of motion of the heart's surface by minimizing an energy function based on a LSTM-RNN architecture, is then used to learn the relationship between the applied force, and to find accurate mapping between the two. Our proposed solution is associated with force sensing devices, such as biocompatibility and integration.

Index Terms—Force Estimation, deep networks, visual deformation, computer-assisted surgery

1 INTRODUCTION

ROBOTIC-Assisted Minimally Invasive Surgery (RAMIS) emerged from the need to address some deficiencies associated with traditional Minimally Invasive Surgery (MIS) and open procedures [1]. The revolutionary technologies utilized by RAMIS systems provide motion scaling and tremor filtering which stabilize the instruments and improve surgery precision [1], [2]. Furthermore, the added degrees of freedom in the tool tip enhances surgeons' dexterity and results in better clinical outcomes [2]. The small incisions used in RAMIS allow reducing the amount of blood loss during surgery, minimizing trauma to the body, and improving cosmetic results. Patients who undergo RAMIS experience less post-operative pain, faster recovery, and lower mortality and morbidity events [3], [4].

Despite all the benefits offered by RAMIS, current commercially available systems suffer from one major limitation which is the lack of force feedback [4], [5]. This feature is of huge importance since it increases surgeon-patient transparency [6] and allows more natural interaction with delicate tissues, as in the case of the heart (Fig. 1). Without force feedback information, surgeons have no means of knowing how much force is applied to the tissue, which could complicate the surgical task, increase its completion time and, what is worst, result in irreversible injuries [7], [8]. Furthermore, dealing with the absence of this primary

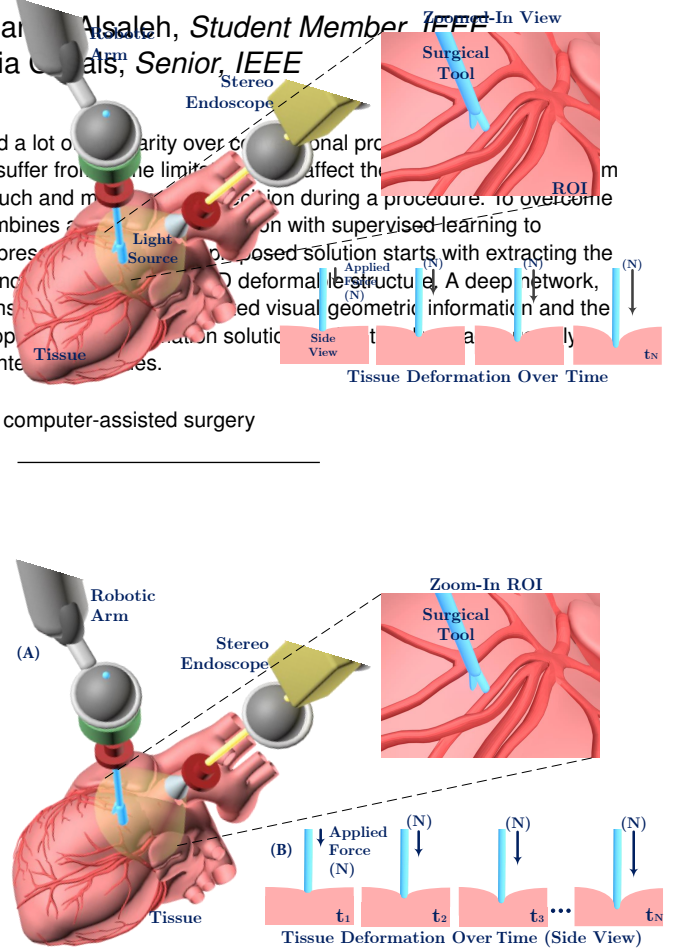


Fig. 1. (A) shows tool-tissue interaction during Robotic-Assisted Surgery which lack force feedback that informs the surgeon about how much force is applied. (B) shows the observable displacements after applying a force, which we obtain using a sensorless approach that relies, in part, on computing the 3D shape of the tissue over time.

sense of touch creates a higher mental workout for surgeons and might be a hazardous source of distraction [9]. For these reasons, numerous researchers have dedicated significant efforts to address the problem of force feedback. However, up to date it is still considered an open problem [10].

In the search for solutions for the lack of force feedback, some researchers have focused their efforts toward developing force sensing devices (FSDs) [11], [12], [13]. These devices can be placed either inside or outside the patient's body. When placed outside, the devices are attached to the robot or its instruments and offer indirect sensing. With this option, the devices measure not only the instrument-tissue interaction forces but also irrelevant force data given by the external/internal surgical environment. Removal of these

- A.I. Aviles is with the Center of Research of Biomedical Engineering Universitat Politècnica de Catalunya, Spain. E-mail: angelica.ivone.aviles@upc.edu
- S.M. Alsaleh and J.K. Hahn are with the Institute for Biomedical Engineering, Department of Computer Science, George Washington University, Washington DC, U.S.A. E-mail: sm57@gwu.edu
- A. Casals is with the Research of Biomedical Engineering, Universitat Politècnica de Catalunya and with the Institute for Bioengineering of Catalonia Spain,, 08028, Barcelona, Spain. E-mail: alicia.casals@upc.edu

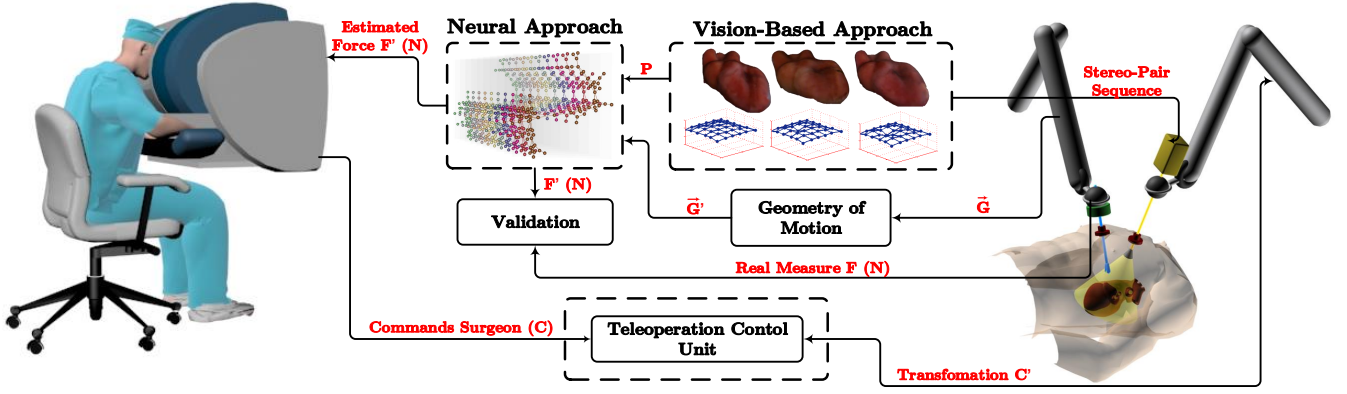


Fig. 2. Flowchart of our approach for estimating applied forces in surgical robotic systems. We first propose a visual approach based on minimizing an energy functional to compute the deformation structure over time. Then, deformation information, together with the geometry of motion, are used as input to an artificial neural network architecture which accurately estimates the applied force.

undesirable measurements is not possible due to hysteresis and because they greatly depend on ambiguous starting conditions [14].

Alternatively, FSDs can offer direct sensing if they are placed close or on the tip of the instrument inside the patient's body. However, the internal location of the sensor introduces numerous problems, including: biocompatibility and sterilization constraints; long-term stability; adaption to surgical tool; size and high cost [15], [16]. All these limitations put severe restrictions to the adoption of FSDs in real surgical environments. Therefore, the use of FSDs can be replaced by Force Sensory Substitution (FSS) in which physical properties of the environment are sensed using an alternate sensing modality. In our case, estimating force information through an alternative sensory channel can substitute the use of FSDs. The potential benefits of FSS for force feedback in teleoperation tasks were first explored by Massimino and Sheridan [17]. In that work, they demonstrated that the use of sensory substitution offers a significant advantage over procedures without force feedback, and eliminates instabilities often associated with FSDs.

When force sensory substitution is used, force feedback can be transmitted to the surgeon through other sensory modalities, specifically auditory, vibrotactile or visual feedback, which provide the surgeon with a representation of the forces he/she is applying with the robotic telemanipulators. In the case of auditory, changes in force are represented as sound signals, like beeps for example. However, some studies have demonstrated that additional sounds during procedures, coming from the medical team or monitors in the operating room, are unavoidable and may interfere with the auditory feedback [7]. Additionally, continuous sounds during long procedures can be a source of discomfort and/or annoyance to the surgeon.

Alternatively, vibrotactile sensory substitution approaches rely on vibration cues to convey information about touched objects. According to the literature, different works have been reported using this sensory substitution. Some examples can be seen in [18]. From one side, they proved

that vibration information is preferred over not having any feedback, in most of the cases, over short periods of time. Nonetheless, vibrotactile feedback offers limited information as it is complicated to convey both force direction and magnitude at the same time due to the high-frequency fluctuations in force and displacement [19], [20], [21]. Moreover, vibrotactile feedback has the drawback of being uncomfortable for the surgeon if used for a long period of time [22].

Visual feedback on the other hand is considered a promising sensory substitution suitable for clinical adoption [14]. With this alternative, surgeons perceive force information via visual cues of tool-tissue interaction. Various studies have investigated the feasibility of visual feedback on conveying force information for surgeons while performing delicate tasks. Investigation results show improved performance among novice surgeons and decreased inconsistencies [14], [23]. Out of the different FSS modalities, in this work we chose visual feedback as it has proven to offer more advantages than other alternatives. Moreover, the use of visual information has been proven to be very reliable for force estimation as all RAMIS settings include a videoscopic view of the operation. Thereby, in order to avoid using force sensors, we can employ the available visual information of the tool-tissue interaction and relate it to the applied force.

In this paper we propose a new approach to estimate the applied forces during RAMIS interventions (see Fig. 2). Our approach uses visual-geometric information, in which tissue deformation is computed by minimizing an energy function while geometric data are provided by the robot. Then, this information is used in a learning system that finds the nonlinear relationship between the given data and use it to estimate the applied force. The proposed approach has the benefits of eliminating the need of add-on sensors which require biocompatibility studies and knowing in advance the physical properties of the deformable tissue.

1.1 Related Work

Vision-based force estimation can incorporate explicit knowledge of the mechanical properties of the tissues. How-

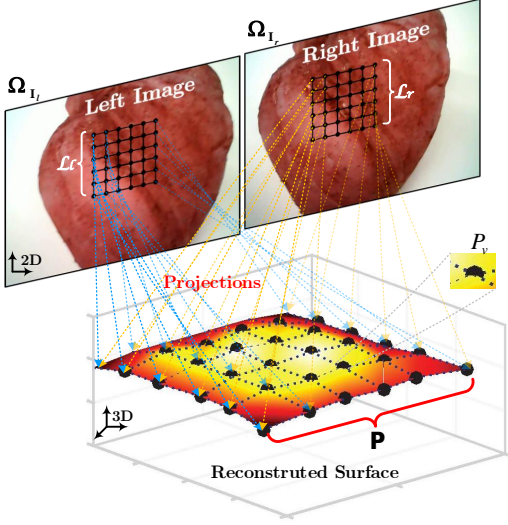


Fig. 3. The 3-dimensional tissue surface is reconstructed from the projections of homologous points on the left and right lattices defined for each stereo-pair image.

ever, this requires both complex calculation and adaptation to each tissue. To avoid these drawbacks, knowledge about the tissue properties can be learned implicitly from the data. This makes the system more suitable for real-time solutions since the learning can be optimized for faster computation.

The viability of using visual information to estimate the applied forces has been demonstrated in different scenarios. In 2D, Greminger et al. in [24] used Dirichlet to Neumann map to estimate the force distribution applied to a deformable object for microassembly and biomanipulation. Authors measured the displacement field of the contour of the object and then used a template matching based on linear elasticity equations. Similarly, authors in [25] modeled the deformation by introducing contour information of the object, together with its mechanical properties, into the boundary element method. Then, deformation data was used to compute the applied forces by means of a capacitance matrix. The disadvantage of this proposal is the need of a prior knowledge of the object's material properties.

The concept of virtual template for computing the deformation of the object, using monocular images, was presented in [26]. In that work, authors assumed that the surface of the object is a smooth function with local deformation. Then, they used a strain-stress relation together with the penetration depth to estimate the force. Authors in [27] applied a mesh-based model to characterize the deformation based on stereo-endoscopic images. Afterwards, a spring-damper system was used to compute interaction forces. Authors in [28] attempted to improve the realism of visual and haptic feedback in a cell injection system by using a 3D nonlinear mass-spring-damper model. The model parameters were identified using offline Finite Element Method (FEM) simulations and the biomembrane geometry deformation was reconstructed using snakes based visual tracking. However, as shown in [29], the use of mass-spring models offers limited accuracy, and the FEM-based parameters computation requires additional modeling efforts.

More recently, some researchers have investigated the use of soft computing to improve the accuracy of the force

estimation. Authors in [30] computed the applied force using a 2-layers feedforward network incorporated into a deformable template matching algorithm. The deformable template was an iterative computation of the object's edge, using Canny's method. Karimirad et al. in [31] used a feedforward Artificial Neural Network to estimate the force applied to cells during micromanipulation. The neural network was trained on geometric features of the cells, including deformation, orientation, and size. These features were extracted using various image processing techniques under different known force conditions. Two different hybrid intelligent systems were proposed in [32] to model the tool-tissue force in laparoscopic surgery: an adaptive coevolutionary fuzzy inference system, and an adaptive neuro-fuzzy inference system. Both systems were trained on three different geometric features extracted from a 2D simulated deformable model: angle and depth of maximum deformation and width of displacement constraint. Nonetheless, experiments in both works, [31] and [32], were only conducted in 2D space.

In our previous work [33], an energy minimization strategy was applied to compute a deformation structure from the acquired stereo image sequences. The deformation structure, along with geometric data from the robotic manipulator, was used as an input to a Recurrent Neural Network (RNN) which was trained using the adapted Levenberg-Marquardt method. A modification of the RNN architecture was presented in [34], in which three main states: local delay, global delay, and no delay, were defined. With the aim of increasing previous system accuracy, in a recent work [35] we used a Long-Short Term Memory RNN (LSTM-RNN) architecture. This LSTM-RNN allowed preserving information for a longer period of time, which enforced constant error flow.

Unlike our previous works, in this paper we introduce a modification of our vision approach which takes into account complex behavior of the deformation structure. Moreover, we reduce complexity and accelerate the performance of our deep network by doing dimensionality reduction to keep only features relevant to force. We also introduce some modifications to the network architecture in order to stabilize force estimation and find the best trade-off between accuracy and computation time.

2 NEURO-RECURRENT-VISION APPROACH

The force estimation strategy proposed in this work is part of the robotic surgical system shown in Fig. 2. In a general RAMIS setting, a surgeon controls the robotic manipulator through a teleoperation control unit that scales and transforms the given commands into relative motion. A stereo pair camera is used to track this motion and feed the image sequences to the vision-based module, which is the first part of our estimation strategy. This module uses the acquired visual information to retrieve the deformation observed on the tissue surface after applying a force. The structure deformation information, along with the geometry of motion given by the robotic manipulator, are given as input to the second module, which is the neural approach. In this module, a Long-Short Term Memory - Recurrent Neural

Network (LSTM-RNN) is trained to analyze the given information and map it into an accurate force estimation. As a final step, the estimated force is validated against the real force measurement given by a robotic sensor attached to the surgical tool for training and validation purposes.

In the remainder of this section our approach is explained in details. Subsection 2.1 describes how a 3D deformation map is reconstructed from the visual information using an energy minimization functional. This is followed by a description of the supervised recurrent learning method that constructs an accurate estimation model that maps both the visual and geometric information into force.

2.1 3D Deformable Shape Recovery

The first part of our solution for estimating the applied force is the computation of the deformation structure, as shown in Fig. 2. In this work, 3D shape recovery is accomplished by minimizing an energy functional reformulated using the l2-regularized optimization class. Moreover, in order to reduce the computational time, we parametrize the changes produced on the tissue surface using a set of linearly independent vectors. In the remainder of this section, we present a formulation that allows recovering the deformation produced when a force is applied on the tissue surface over time.

Let us assume that $\mathbf{I}_l^t : \Omega_{\mathbf{I}_l} \rightarrow \mathbb{R}^2$ and $\mathbf{I}_r^t : \Omega_{\mathbf{I}_r} \rightarrow \mathbb{R}^2$ are the left and right image views from a stereo pair image acquired at each instant time t , where $\Omega_{\mathbf{I}_l}$ and $\Omega_{\mathbf{I}_r}$ are their corresponding domains. Since during a medical procedure the surgeon is interested only in the region to be repaired, for example a vein, computational cost can be reduced by defining a region of interest (ROI).

Definition 1. A m -dimensional lattice is the \mathbb{Z} -linear span of a set of k linearly independent vectors in \mathbb{R}^m .

Since human beings perceive the world as a 3D representation, and taking Definition 1, we handle the specified ROI using a 3D lattice. Let $\mathcal{L}_l : \Omega'_{\mathbf{I}_l} \subseteq \Omega_{\mathbf{I}_l}$ and $\mathcal{L}_r : \Omega'_{\mathbf{I}_r} \subseteq \Omega_{\mathbf{I}_r}$ be 2D lattices defined at each image view respectively. Then, the 3D lattice is computed from the projections of the corresponding lattice points on $\hat{\mathbf{I}}_l \subseteq \mathbf{I}_l$ and $\hat{\mathbf{I}}_r \subseteq \mathbf{I}_r$ (see Fig. 3). Let \mathbf{P} be the result of such correspondences and v be the number of lattice points, $P_v \in \mathbf{P}$ where $P_v = (y_1, \dots, y_m) \in \mathbb{R}^m$. Considering that initially lattice points are evenly spaced, then the changes produced on the tissue surface, over time, are computed by minimizing the total energy, \mathbf{E}_t , such that the optimal \mathbf{P} can be found using the following equation:

$$\mathbf{E}_t(\mathbf{P}) = \mathbf{E}_\Phi(\hat{\mathbf{I}}_l^t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}), \hat{\mathbf{I}}_r^t(\mathbf{x})) + \gamma \mathbf{E}_\Psi(\Gamma(\mathbf{x}; \mathbf{P})) + \mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P}) \quad (1)$$

where \mathbf{E}_Φ is the discrepancy measure term, \mathbf{E}_Ψ denotes the penalization term used to obtain a plausible transformation, $\gamma \in \mathbb{R}^+$ is the parameter that controls the quality of the data fit, \mathbf{E}_Λ gives a constraint to preserve shape, \mathbf{x} is a vector containing the coordinates, and Γ is the deformation model.

The deformation model is an essential factor that defines how fast and accurate the approach is. In order to find a compromise between computational cost and accuracy,

we characterize the lattice points using the tensor product of b-splines as they demand low running time, allow multiresolution, have optimal mathematical properties and keep affine invariance [36]. Moreover, and despite that there are different deformation models that are widely used in medical applications, b-splines have the additional advantage of being able to handle complex deformations. The deformation model is defined as follows:

Definition 2. Let Γ be the characterization of \mathbf{P} over time and consider $\xi(\cdot)$ as the basis spline function of degree n at d -dimensional space. Then, the lattice points behaviour is given by:

$$\Gamma(\mathbf{x}; \mathbf{P}) = \sum_{j_1=0}^n \dots \sum_{j_d=0}^n \overbrace{\mathbf{P}_{j_1, \dots, j_d}}^{\text{control points}} \underbrace{\prod_{k=1}^d \xi_k(\mathbf{x}_k)}_{\text{tensor product}} \quad (2)$$

In particular, we use the cubic basis splines expressed as:

$$\begin{aligned} \xi_0(\mathbf{x}) &= (1 - \mathbf{x})^3/6 & \xi_1(\mathbf{x}) &= (4 + 3\mathbf{x}^3 - 6\mathbf{x}^2)/6 \\ \xi_2(\mathbf{x}) &= (1 - 3\mathbf{x}^3 - 3\mathbf{x}^2 + 3\mathbf{x})/6 & \xi_3(\mathbf{x}) &= \mathbf{x}^3/6 \end{aligned} \quad (3)$$

In order to achieve a robust and efficient solution, in this work, we use the l2-regularized optimization class ($\|\cdot\|_{L_2(\Omega')}$). We selected this class because it offers many advantages including: it is differentiable, strictly convex, offers a unique solution, and works well with smooth data [37].

We now turn to reformulate the energy functional defined in Eq. 1. The discrepancy term, \mathbf{E}_Φ , is computed using the sum of squared differences method. This was selected because it has a low computational cost and offers an optimal result when images are acquired with the same sensor, as it is in our case. But due to its sensitivity to changes of intensity, we carried out a normalization process. Moreover, since this is an ill-posed problem, in the sense of Hadamard [38], it is necessary to have a penalization term to restrict the solution space and impose stability to the energy functional defined in Eq. 1. Thus, to obtain a well-posed problem we rewrite the penalization term \mathbf{E}_Ψ using Tikhonov regularizer. Finally, we add a soft constraint for volume-preserving mappings (incompressibility). Let S be the number of overlapping pixels. Taking previous statements, Eq. 1 results in:

$$\begin{aligned} \hat{\mathbf{E}}_t(\mathbf{P}) &= \underbrace{\|\mathbf{E}_\Phi\|_{L_2}^2}_{\text{discrepancy}} + \underbrace{\gamma \|\mathbf{E}_\Psi\|_{L_2}^2}_{\text{penalization}} + \underbrace{\|\mathbf{E}_\Lambda\|_{L_2}^2}_{\text{constraint}} \\ &= \frac{1}{S} \left(\int_{\mathbf{x} \in \Omega'} \|\hat{\mathbf{I}}_l^t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}) - \hat{\mathbf{I}}_r^t(\mathbf{x})\|^2 d\mathbf{x} \right. \\ &\quad \left. + \gamma \sum_{i=1}^d \int_{\mathbf{x} \in \Omega'} \|\nabla \Gamma_i(\mathbf{x}; \mathbf{P})\|^2 d\mathbf{x} + \int_{\mathbf{x} \in \Omega'} \|\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P})\|^2 d\mathbf{x} \right) \end{aligned} \quad (4)$$

In search of practicality and efficiency, we use a discretize-then-optimize process. Strictly speaking, after defining the continuous optimal energy functional, we transform it into a standard optimization problem by discretizing the original one. Now, in order to have a standard

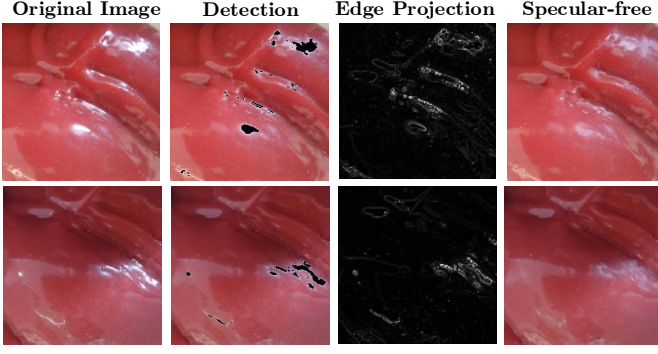


Fig. 4. Specular highlights cause major tracking disturbance. We deal with this issue using a real-time detection and inpainting approach that accurately recovers a specular-free image.

optimization problem, discretization of Eq. 4 is expressed as:

$$\hat{\mathbf{E}}_t(\mathbf{P}) = \frac{1}{S} \left(\sum_{\mathbf{x} \in \Omega'} \|\hat{\mathbf{I}}_t^t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}) - \hat{\mathbf{I}}_r^t(\mathbf{x})\|^2 + \gamma \sum_{i=1}^d \sum_{\mathbf{x} \in \Omega'} \|\nabla \Gamma_i(\mathbf{x}; \mathbf{P})\|^2 + \sum_{\mathbf{x} \in \Omega'} \|\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P})\|^2 \right) \quad (5)$$

where the soft constraint $\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P}) = \det(\nabla \Gamma(\mathbf{x}; \mathbf{P}))$.

2.1.1 Robust 3D Shape Recovery

During the process of recovering the temporal 3D deformable structure, different factors can affect the performance of the visual approach. Therefore, in order to increase robustness of our solution, we handle potential sources of error as we want to accurately reconstruct the deformable surface. One source of error that might affect the reconstruction precision is the specular highlight regions that appear on the surface of the heart. These bright spots appear on surfaces with high reflectivity and occlude the underlying visual information causing uncertainty in the tracked ROI. To eliminate this artifact, we integrate our previous work related to detection and removal of specular highlights in which a hybrid detection technique, based on saturation and intensity color attributes and Wavelet Transform Modulus Maxima, was used to detect the affected regions accurately. We then correct the detected regions using a dynamic search based inpainting that smoothly propagates pixel information from the surrounding area (see Fig. 4). The solution was optimized to process the acquired images in real time (for more details see [39]).

Another potential source of error when tracking the surface deformation is the partial occlusion of the tracked region of interest (ROI). Occlusion makes the tracking process more challenging and could cause tracking failure as the algorithm will not have enough information about the occluded part of the surface. In RAMIS settings, the tracked ROI may be partially occluded for a short period of time, by a surgical tool or blood, which might hide useful information about the surface and affect the tracking precision (see Fig. 5). This source of error needs to be eliminated as precision is an essential factor in medical applications.

The energy functional given in Eqs. (4) and (5) so far does not reduce the influence of outliers, i.e. an error can be

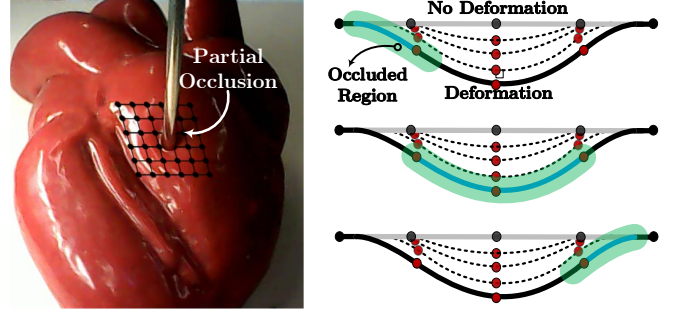


Fig. 5. Surgical tool can partially occlude the tracked region of interest which affects the 3D shape recovery over time (Left side). Right side shows a side view of occluded lattice regions from different viewpoints.

produced when a partial-occlusion appears. In order to deal with this issue, we reformulate the total energy including a least-squares estimator, called M-estimator, defined as:

Definition 3. The M-estimation process substitutes the minimization of $\sum_i \mathbf{r}_i^2$, where \mathbf{r} is the residual error, with $\sum_i \rho(\mathbf{r}_i)$, where ρ is a specific function called M-estimator.

Thus, the idea is to include an M-estimator in our energy functional, described in Eq. 5, for penalizing the largest residual errors. The new total energy, $\tilde{\mathbf{E}}_t$, is expressed as:

$$\tilde{\mathbf{E}}_t(\mathbf{P}) = \frac{1}{S} \left(\sum_{\mathbf{x} \in \Omega'} \rho(\hat{\mathbf{I}}_t^t(\Gamma(\mathbf{x}; \mathbf{P}) + \mathbf{x}) - \hat{\mathbf{I}}_r^t(\mathbf{x})) + \gamma \sum_{i=1}^d \sum_{\mathbf{x} \in \Omega'} (\Gamma_i(\mathbf{x}; \mathbf{P}))^2 + \sum_{\mathbf{x} \in \Omega'} (\mathbf{E}_\Lambda(\mathbf{x}; \mathbf{P}))^2 \right) \quad (6)$$

We use the robust Huber's M-estimator [40] for ρ in which c is a positive tuning constant and ρ is expressed as:

$$\rho_{huber}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq c \\ c|x| - \frac{1}{2}c^2 & \text{otherwise} \end{cases} \quad (7)$$

Once the total energy, $\tilde{\mathbf{E}}_t$, is defined, we turn to finding the optimal value, \mathbf{P} , at each time instant t . To do that, we use the Levenberg-Marquardt (LM) method [41], [42]. LM combines the stability of the gradient descent and the fast convergence of the Gauss-Newton. LM makes use of a damping parameter, δ , in order to switch between the gradient descent and the Gauss-Newton. When δ is small, it acts as Gauss-Newton with the difference that it uses a trust-region with radius Δ_h instead of a line search. While δ is large, it performs as gradient descent. So, search direction, d_h , at iteration h is computed as follows:

$$\begin{aligned} (\mathbf{J}_h^T \mathbf{J}_h + \delta \mathbf{I}) d_h &= -\mathbf{J}_h^T \mathbf{r}_h \quad \delta > 0 \\ d_h(\delta) &= -(\mathbf{J}_h^T \mathbf{J}_h + \delta \mathbf{I})^{-1} \mathbf{J}_h^T \mathbf{r}_h \end{aligned} \quad (8)$$

where \mathbf{J} is the Jacobian, \mathbf{r} is the residual vector, and \mathbf{I} is the identity matrix.

Observation 1. The optimization process was restricted with a maximum of 25 iterations, achieving a trade-off between accuracy and computational speed.

It is worth mentioning that the lattice used for the 3D shape recovery is as well used for knowing when the

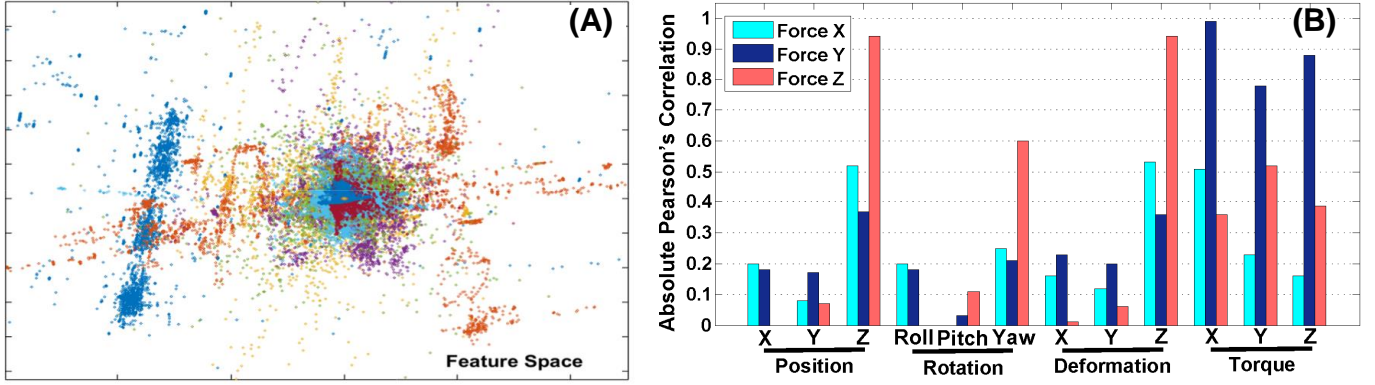


Fig. 6. Part A shows all the variables that can be extracted during a robotic-assisted surgery (total of 50). Nonetheless, only a subset of this large feature space is relevant and it needs to be identified to improve the supervised learning process. To achieve dimensionality reduction, we used Pearson correlation test and plot B shows the subset of the most correlated features. We can see that the most dominant ones are in the Z direction because in the sample dataset that we picked to illustrate the correlation results, the force was mainly applied in that direction.

robotic manipulator gets in-touch with the tissue. For this, we used as a reference the control points defined in the lattice and start data acquisition when we measure the first displacement at any of those points.

2.2 Force Estimation Strategy

Once the 3D deformable shape is accurately recovered, the second module in our proposed solution for force estimation is a deep network based on an LSTM-RNN architecture. Our goal is to utilize the deformation information that we extracted from the recovered 3D deformable structure to estimate the force applied to the tissue surface. We first start by describing the dimensionality reduction approach to reduce the complexity of the model. This is followed by a detailed description of the design of the long-short term memory recurrent network architecture that we utilize to map the visual and geometric information into an accurate force measure.

2.2.1 Reduction of Dimensionality

In this work, we use the geometry of motion of the robotic manipulator along with the structure deformation in all directions to train our Neural Network architecture. This gives us a complex and large set of input data to work with, which directly affects the training time and complexity of the NN. One common pre-processing approach that can help reducing the NN training time is dimensionality reduction, which attempts to eliminate redundant and insignificant information by selecting a subset of relevant features for use in model construction. Besides speeding up training, reducing the dimensionality of the dataset before presenting it to the neural network can improve its performance and prevent overfitting which is a common problem that occurs during neural networks training. Overfitting happens when we have an excessively complex model with too many parameters relative to observations. By reducing dataset complexity, the NN will avoid learning insignificant details and only valuable information will be used for training. Moreover, by avoiding overfitting, the learned NN can be generalized to new datasets and have a better predictive performance.

Many dimensionality reduction techniques are available and capable of efficiently reducing the set of variables without a significant loss of information. One common technique is the correlation-based feature selection which evaluates subsets of features to determine the strength of the relationship of each feature with the response variable, in our case the applied force. In this work, we use Pearson product-moment correlation coefficient, ρ , to measure the linear correlation between each of the input features and the applied force. The Pearson correlation coefficient analyzes how much the variation in one variable affects the variation in another and is defined as:

$$\rho^{X,Y} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2} \sqrt{\sum(Y - \bar{Y})^2}} \quad (9)$$

where X represents the input features, Y is the force measure, and \bar{X} and \bar{Y} are their mean values. The correlation coefficient ρ is between -1 and 1 where a value of 1 or -1 means perfect positive or negative correlation respectively. Out of the many datasets we collected during our experimentation, we examined one in which force was mainly applied in the Z direction. Fig. 6-(A) shows the complete feature space, a total of 50 features, while Fig. 6-(B) shows the features with the highest correlation to the force. The position and rotation come from the geometry of motion while the deformation is from the 3D surface reconstruction using our proposed approach. We also show the torque that is acquired as used as ground truth. The plot shows the absolute value of Eq. 9 with coefficients ≥ 0.4 . This observation indicates that it is feasible to reduce the dimensionality of the feature space to include only those with high correlations to force.

2.2.2 Supervised Recurrent Learning

In this subsection, we describe our soft computing based force estimation strategy. Particularly, we use Artificial Neural Networks (ANNs) to find the relationship between the inputs (visual and geometric information) and the force, more specifically Recurrent Neural Network (RNN). Unlike conventional networks, the layers in RNNs introduce memory by having feedback connections between their units, which enables dynamic temporal processing instead of a hierarchical one. Moreover, RNNs can be viewed as

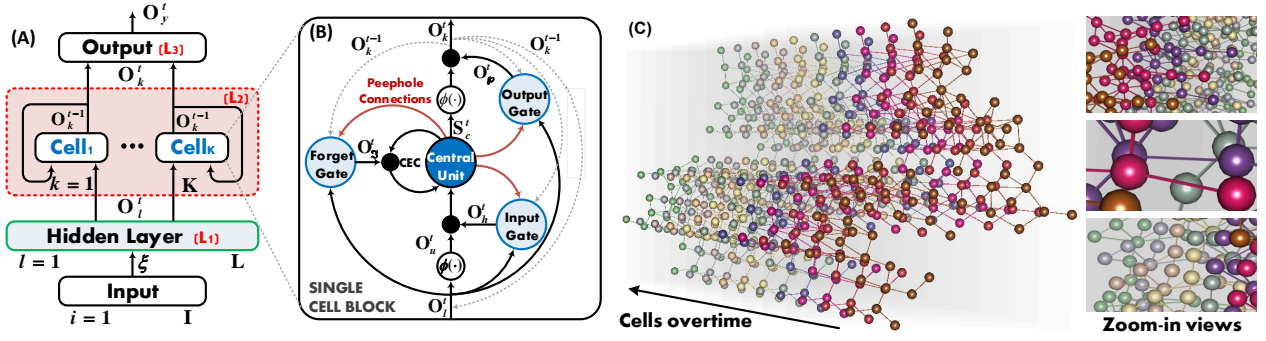


Fig. 7. In order to estimate the applied force, we used an architecture based on LSTM-RNN (part A) which combines basic units with cells. Part B shows a single cell block in detail and shows that each of the cells is composed of a set of units that enforce constant error flow which helps stabilizing force estimation over time. Additionally, part C shows an illustration of the hidden layer with 10 cells over time.

deep networks when folded out in time [43], and have the advantages of handling noise-contaminated data and creating complex input-output relationship. Nevertheless, the vanishing gradients problem, where error-signals exhibit exponential decay as they are back-propagated through time, has a direct impact on the performance of RNNs [44]. In this work, we utilize LSTM to overcome this problem and improve the accuracy of force estimation in RAMIS.

LSTM are specially designed to store and retrieve information over long periods of time and enforce constant error flow by using specialized units, called cells. An LSTM layer has one or more recurrently connected memory cells composed of a central unit, and specialized input, output, and forget gates. The input and output gates are multiplicative units that protect the memory content from perturbations. On the other hand, the forget gates release irrelevant information by resetting the memory cell when the information stored there is not useful anymore [44]. These three gates have access to the central unit through peephole connections. The learning process is defined as:

Definition 4. Given a set of \mathfrak{N} training samples in the form of input-output pairs $\{(x_1, y_1), \dots, (x_{\mathfrak{N}}, y_{\mathfrak{N}})\}$, where x is a feature vector and y is its corresponding target value, supervised learning finds a function $f : X \rightarrow Y$ that maps the input space, X , to the output space, Y , and works well on unseen inputs x .

Our architecture, as illustrated in Fig. 7, is composed of two types of hidden layers: with basic units (Layer 1) and memory cells (Layer 2). Following the notation presented in Fig. 7-(A), let ξ be the input vector with I inputs, L the number of units, and K the number of cells with C memory cells in each block. Let w and b denote the weights and the bias respectively, and ϕ the activation function, which in this case is the log-sigmoid function (a.k.a logistic function). Then, the outputs, O_l^t , are computed as follows:

$$O_l^t = \phi(\sum_{i=1}^I w_{il} \xi_i^t + b_l) \text{ for } l = 1, \dots, L \quad (10)$$

Each output of Layer 2, O_k^t , is defined by the relation of a set of units, as depicted in Fig. 7-(B). Consider h , φ and \mathfrak{S} as the input, output and forget gates, their corresponding outputs are defined as:

$$\begin{aligned} O_h^t &= \phi(\sum_{l=1}^L w_{lh} O_l^t + \sum_{k=1}^K w_{kh} O_k^{t-1} + \sum_{c=1}^C w_{ch} S_c^{t-1}) \\ O_{\varphi}^t &= \phi(\sum_{l=1}^L w_{l\varphi} O_l^t + \sum_{k=1}^K w_{k\varphi} O_k^{t-1} + \sum_{c=1}^C w_{c\varphi} S_c^t) \\ O_{\mathfrak{S}}^t &= \phi(\sum_{l=1}^L w_{l\mathfrak{S}} O_l^t + \sum_{k=1}^K w_{k\mathfrak{S}} O_k^{t-1} + \sum_{c=1}^C w_{c\mathfrak{S}} S_c^{t-1}) \end{aligned} \quad (11)$$

Continuing with the notation presented in Fig. 7-(B), the output of the unit O_u^t and the memory cell state S_c^t are obtained as follows:

$$\begin{aligned} O_u^t &= \phi(\sum_{l=1}^L w_{lh} O_l^t + \sum_{k=1}^K w_{ku} O_k^{t-1}) \\ S_c^t &= O_{\mathfrak{S}}^t S_c^{t-1} + O_h^t O_u^t \end{aligned} \quad (12)$$

using Eqs. 11 and 12, we can describe the output of each cell as:

$$O_k^t = O_{\varphi}^t \phi(S_c^t) \text{ for } k = 1, \dots, K \quad (13)$$

Finally, the output of the network, O_y^t , is given by:

$$O_y^t = \phi(\sum_{k=1}^K w_{ky} O_k^t + b_y) \text{ for } y = 1, 2, 3 \quad (14)$$

where the output units are the force in X , Y and Z directions. Using this LSTM-RNN based architecture, the vanishing gradient problem is solved. Thus, a gradient-based algorithm can be used. In this work, we apply Backpropagation Through Time (BPTT), which unfolds the network over time by replicating the network and sharing the weights.

Observation 2. Notice that the learning process is carried out offline since its goal is to find the optimal parameters. Once the adjusting parameters (weights) are found, they are used in our system in real-time.

Taking into account observation 2, the training process for the deep network (Fig. 7-(C)) is performed once to learn the optimal parameters and the mapping function from the input to the output space. Once the network is trained, it can be used for estimating the applied forces in real time systems without the need of carrying out the training process for each subject. This is based on the fact that, according to [45], [46], generic material properties of the human heart tissue can be modeled and, in consequence, these properties can be learned using the LSTM-RNN architecture then generalize the model across subjects. The proposed approach can be generalized to handle other tissues, requiring only a

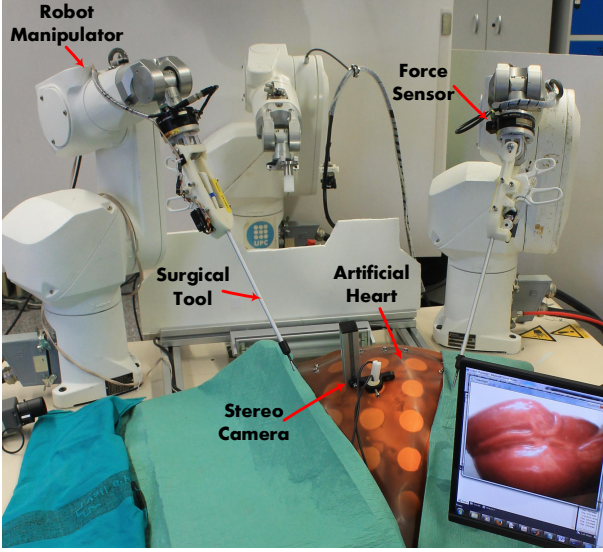


Fig. 8. The realistic surgical setting, with typical RAMIS surgical setup, used to obtain the two ex-vivo datasets. The force sensor is used to obtain the ground truth to validate our estimation.

single training run in order to obtain the mapping function for the new tissue. The user of the system can then select the desired function during the real time procedure or can even choose to train a function that can handle different tissue types.

3 EXPERIMENTAL RESULTS

This section describes in detail the experimentations that we conducted to validate the accuracy of the proposed solution.

3.1 Data Description

To evaluate our proposal, we used both in-vivo and ex-vivo datasets (see Fig. 10-(A)).

The in-vivo dataset [47] is from a porcine and exhibits tissue deformation, due to tool interaction, and was used to evaluate our deformation approach. This sequence is composed of stereo-pair images of size 720x288 recording during a period of 450 sec.

Moreover, we acquired the ex-vivo datasets using the experimental setup showed in Fig. 8. It is composed of a stereo camera, a set of robot manipulators (Stäubli RX60B), and an ATI Gamma SI-32-2 force sensor which we used to acquire a ground truth for the applied force in order to compare it with our estimation. We obtained two stereo-pair images sequences of size 640x480 recording during 2100 sec.

As for the ex-vivo datasets, we used two artificial hearts made of ECOFLEX 0030, which has mechanical properties similar to those of human tissues, to imitate variations between two different subjects. From a technical point of view ECOFLEX material allows comparing our approach with other research since it is widely used and considered a standard material for experimentation in clinical environments (e.g. [48], [49], [50]). Moreover, ECOFLEX facilitates the continuous experimentations avoiding at the same time hygienic issues.

It is noteworthy that during the acquisition of our ex-vivo datasets we did not take the dynamics of the heart into consideration. However, it will be included in a future work.

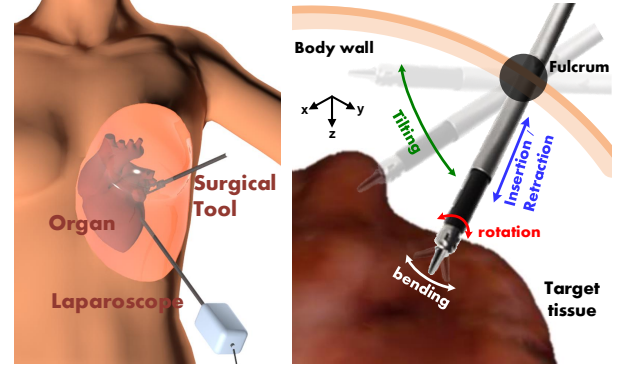


Fig. 9. Illustration of the palpation and exploration surgical tasks used to test the efficiency of our solution.

3.2 Tasks Description

The datasets described in Subsection 3.1 were acquired by doing general inspection through palpation over the tissue and region of interest while varying three main factors over time: position, orientation, and illumination. An illustration of palpation actions can be seen in Fig. 9.

General palpation is necessary during different clinical activities such as tumor detection, tissue cutting, and needle-based procedures; it is an actuation very representative for this study. Palpation is relevant for RAMIS since during procedures, surgeons perform different tasks part of which is to avoid penetration of the tissue and control the applied force.

3.3 Evaluation Scheme

Our evaluation scheme is divided into two parts. The first part uses an in-vivo dataset to evaluate the following:

- Inspection of the displacement field: and Fig. 10-(B);
- Careful comparison of the residual error of our deformation approach: Table 1;
- Visual and numerical prove of the potential of our variational framework with in-vivo datasets: Fig 10-(A)/(B) and Table 1;

In the second part, we used our two ex-vivo datasets with a provided ground truth and performed the following evaluations:

- Visual examination of the displacement field: Fig. 10-(C);
- Convergence of our energy functional (Eq. 4): Fig. 10-(D);
- Comparison between the estimated and real displacement at contact point: Fig. 10-(E);
- Associated strength between the real and estimated force: Fig. 11;
- Statistical analysis of adaptability of our force estimation strategy: Table 2.
- Comparison between the real and estimated forces in the (X,Y,Z) directions: Fig. 12;
- Stability over longer periods of time of our proposal for estimating the interaction forces: Fig. 13;

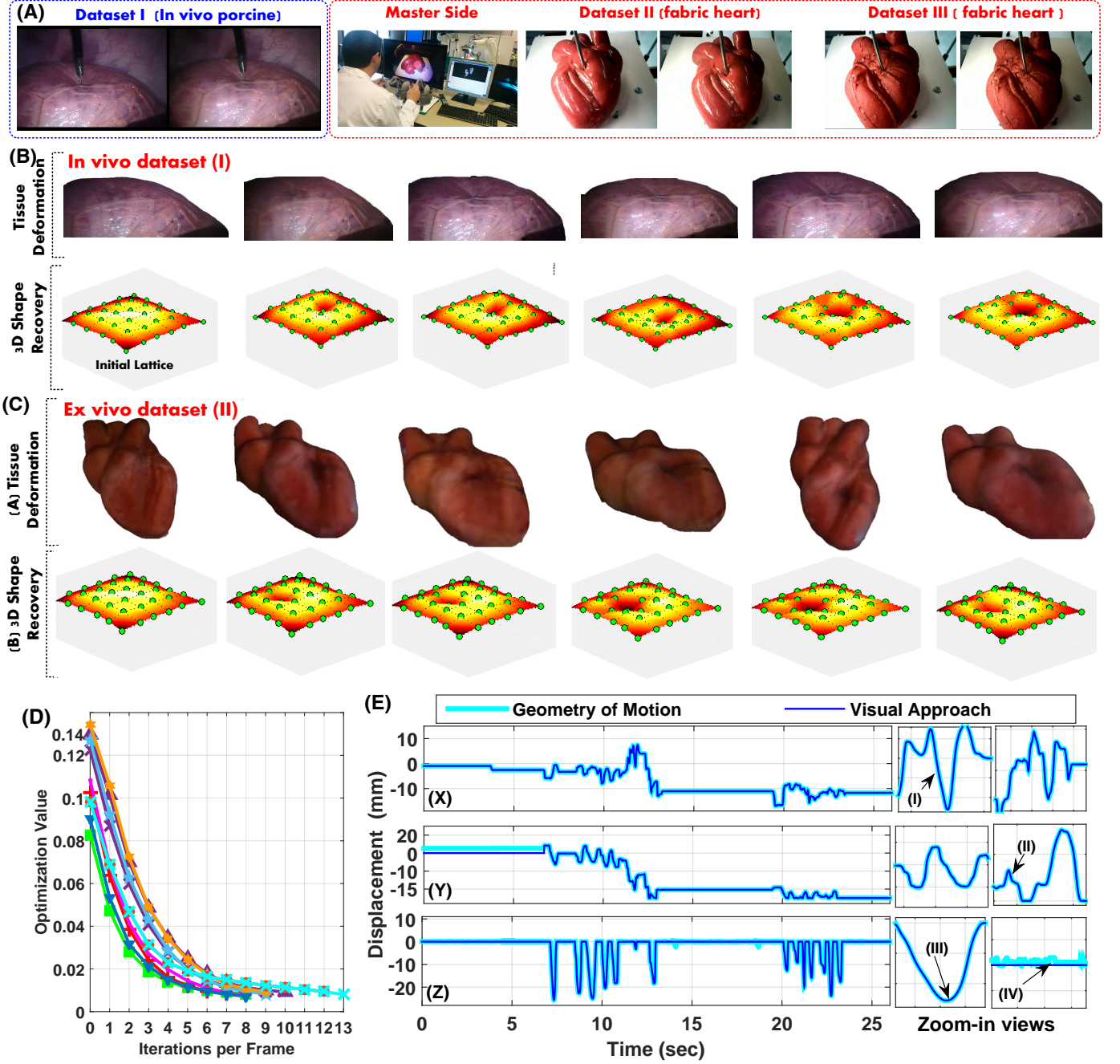


Fig. 10. Part (A) shows the raw data of the three different datasets used to evaluate our proposal (one in-vivo and two ex-vivo). Tissue deformation that result from applying a force at different time instants is illustrated in parts (B) and (C) along with the recovered 3D deformable structure using our proposed visual approach. The convergence plot at part (D) shows the number of iterations it took to get the results at (C). Finally, plots at part (E) show a comparison between the computed displacement (at contact point) in X,Y,Z directions against the reference measurements given by the geometry of motion of the robot from dataset II. The zoom-in views demonstrate the high estimation accuracy of our approach even during complex deformation as it can capture small (I-II) and large displacements (III). It also eliminates the noise in the geometry of motion as shown in (IV).

3.4 Results and Discussion

In order to prove the benefits of our proposal, in this subsection, we offer detailed evaluation of both our visual-based and force estimation approaches.

3.4.1 Visual-based Approach

We evaluated the performance of our visual-based approach using dataset I-II (see Fig. 10-(A)). First, Fig. 10-(B)/(C) show tissue deformation that results from applying force and illustrates the recovered 3D deformation structure, bounded

by our defined grid, over some time instants where darker shades of red represent more intense deformation at contact point. The plots clearly show pleasant visual results of the deformation field with both in-vivo and ex-vivo data.

We then took the results from the in-vivo data, Fig. 10-(B), and offer a quantitative analysis of our energy functional (Eq. 4). The results are reported in Table 1 in which Exp. 1 and Exp. 2 show that without penalization given by the M-estimator and the two regularizers (Eqs. 6-7), the residuum was about 0.1348 and 0.2657 respectively.

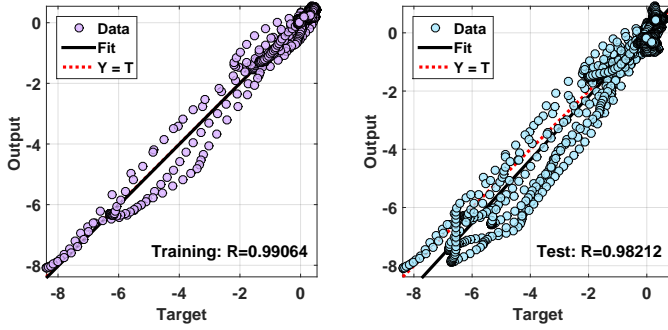


Fig. 11. These linear regression plots show the associated strength between the real (target) and estimated force (output) measurements of both training and test datasets. In both sets, the points fit a line showing a tight relationship between the measurements and demonstrating the accuracy of the force estimation.

TABLE 1
Residual Error evaluation of our deformation approach

Exps.	Energy Functional (Eq. 4)	Minimum
1	E_Φ without ρ_{huber} , $E_\Psi = 0$, $E_\Lambda = 0$	0.2657
2	E_Φ with ρ_{huber} , $E_\Psi = 0$, $E_\Lambda = 0$	0.1348
3	E_Φ with ρ_{huber} , E_Ψ , $E_\Lambda = 0$	$1.7896e^{-03}$
4	E_Φ with ρ_{huber} , E_Ψ , E_Λ	$3.1584e^{-05}$

Comparing that to Exp. 3, we can see that including Tikhonov regularizer resulted in a minima in the order of magnitude 10^{-3} ; while adding a volume preserving term, as in Exp. 4, clearly offered the best minima in the order of magnitude 10^{-5} . With this, we conclude that the combination between the ρ_{huber} with the two regularizers offered a significant difference in order of magnitude.

From the results of Fig. 10-(C), we analyzed both the convergence of our functional and the accuracy of the computed deformation. The plot at Fig. 10-(D) shows that the minimization of our functional, on different frames, needed less than 25 iterations to get the minima. For this reason, we limited the number of iterations according to Observation 1.

Moreover, by acquiring the geometry of motion from the robotic manipulator, we were able to have a ground truth reference at least for the contact point between the tool and the tissue. So in order to evaluate the accuracy of our computed deformation, we compared the displacement value at contact point in X, Y, and Z directions against the reference measurements given by the geometry of motion. The plot at Fig. 10-(E) shows that comparison and the zoom-in views, together with a root-mean-square error (RMSE) smaller than 1mm in all directions, demonstrate the accuracy of our computed measurements even during complex deformation. Furthermore, our visual approach was even able to deal with the mechanical issues that usually exist in the geometry of motion and eliminate the noise as shown in view (IV). This supports the good accuracy and fast convergence of our proposed visual-based approach.

3.4.2 Force Estimation Approach

The ultimate goal of this work is to estimate the applied force in RAMIS scenarios accurately over time. Therefore, we conducted large number of tests to validate our complete

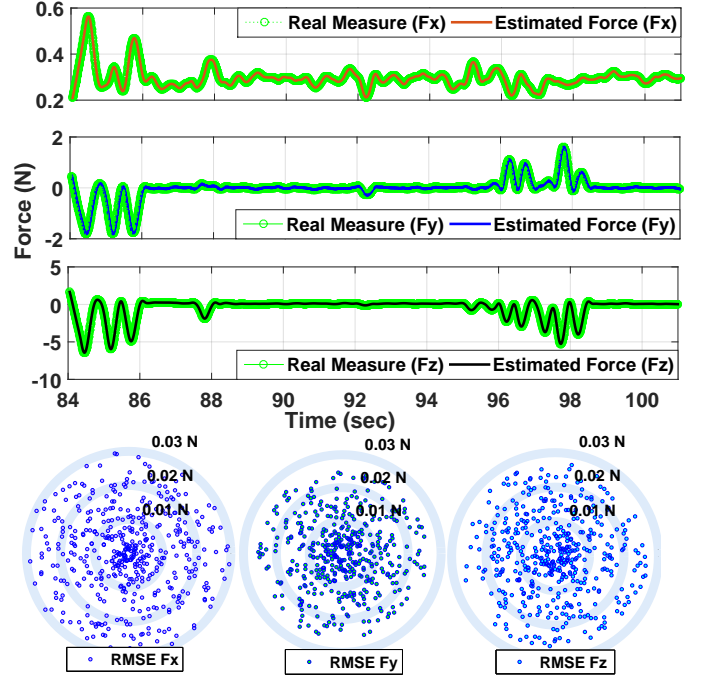


Fig. 12. Plots in top part show the real force measures, in X,Y and Z directions, and those estimated by our approach. Bottom plots illustrate the RMSE results in all directions.

TABLE 2
Statistical nonparametric analysis of our proposal to estimate the applied forces. It takes into consideration the ex-vivo datasets and the real measure.

Input Values	Direction	p-value	Null Hypothesis
Real and Dataset II (ex-vivo data)	x	0.8105	h=0
	y	0.8026	
	z	0.7598	
Real and Dataset III (ex-vivo data)	x	0.8654	h=0
	y	0.8287	
	z	0.8045	

neuro-recurrent-vision solution against the ground truth of the force provided with the ex-vivo datasets.

To validate the accuracy of our LSTM-RNN approach, we tested the associated strength between the estimated and actual forces, and the regression plots in Fig. 11 show that our model is very accurate as there is a strong correlation between the two measures. The red dashed lines in the plots show the ideal solution while the straight black lines are the best linear regression fit between the target and the output. The tight relationship is clear in both training and test dataset as they reported R-values of 0.99 and 0.98 respectively.

It is worth mentioning that our force estimation model was trained once using dataset II, but then the optimal parameters were tested on the two artificial hearts that have slightly different mechanical properties (Dataset II and Dataset III). The results show that our model adapted well to the two dataset and prove that it was able to handle small variations across subjects.

In order to support the previous statement, we ran a statistical analysis on the force estimated on both dataset II

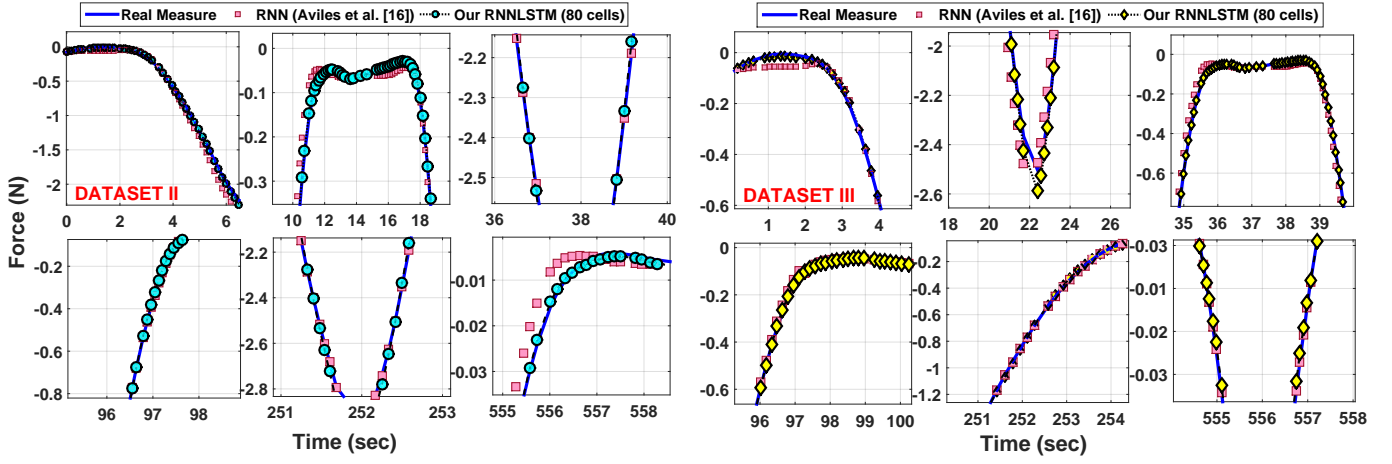


Fig. 13. Stability criteria is shown in these plots using the ex-vivo datasets where the estimated and real force measures are plotted at different time intervals of a longer period of time.

and III and the corresponding ground truth. More specifically, we used the nonparametric Wilcoxon rank sum test to answer the question of whether there is a statistical significant difference between the two estimated forces. The results at Table 2 show that the null hypothesis was not rejected at $p < 0.05$ of significance level. This, together with the big p-value for the three directions (X,Y,Z), led us to conclude that there is no significant difference between the two groups which support our proposal in the sense of adaptability to different subjects.

The accuracy of our solution is further validated by the top plots shown in Fig. 12 in which we compared the estimated force (test data) against the real one in the (X,Y,Z) directions and the results show that the measurements are very close to each other. We also show the RMSE results at the bottom plots in which the error remained less than 0.03N with a concentration of values much lower than 0.02 N in all directions.

Furthermore, we demonstrate the stability of our solution over time by inspecting the results of both datasets at different time intervals (see Fig. 13). The zoom-in views show comparison of the force measures as given by the force sensor, our previous RNN solution [34], and our newly proposed LSTM-RNN solution. As we can see, the added cells with the LSTM architecture improve the accuracy of the results and bring the force measure closer to the actual one. Furthermore, the results we obtained from the LSTM-RNN tend to be more stable, and with no error decay, during long periods of time when compared to those obtained employing only the RNN as it is clear in the last time interval. Based on the results obtained in our evaluation scheme, we report an average RMSE of 0.02 N for all our experiments.

4 CONCLUSION

In RAMIS, surgeons perform delicate procedures remotely through robotic manipulators without directly interacting with the patients. As a result, they lack force feedback that informs them about how much force the surgical tool is applying to the tissue. While force sensing devices are able to provide that information, their size and cost, along

with biocompatibility concerns, prevent them from being fully integrated into the surgical environment. The force sensory substitution approach presented in this work offers a feasible alternative that overcomes the sensors limitations and could potentially improve the precision of RAMIS. The proposed approach combines a computationally efficient visual shape recovery approach with an accurate LSTM-RNN based force estimation model.

By minimizing an optimized energy functional, we were able to recover the 3D deformable structure of the region of interest over time. We ensured the robustness of our shape recovery approach by handling sources of errors and outliers that exist in real surgical environments such as occlusions and uncertainties. Furthermore, we utilized the learning power of deep network by using a LSTM-RNN to relate the extracted visual-geometric information to an accurate force estimation. We obtained a trade-off between computational time and accuracy of our deep network by reducing the complexity of the input space and only considering features with high correlation to force.

The experimental results presented in section 3 verify that vision-based techniques combined with supervised learning provide a feasible and accurate estimate of the applied forces without using force sensors. Experiments included various datasets, in-vivo and ex-vivo, and the computed and estimated results were validated against the ground truth obtained from the robotic manipulator and the force sensor. The advantages of this approach include robustness, accuracy, and stability over long periods of time. This methodology would allow surgeons performing RAMIS to have force feedback and would increase the transparency of interaction with the patient without using force sensors.

Moreover, our solution promises to be useful in robotic-assisted surgery as well as in different situations in which to know the applied force make a difference in the results, including: detection and prevention of diseases or abnormal behaviour (e.g. [51]), needle-based procedures (e.g. [52]), microsurgery (e.g. [53]) and knot tying (e.g. [54]). Thus, this approach can be extrapolated in the above-mentioned situations avoiding in this way the space restrictions, biocompatibility issues and cost of the design new minaturized

force sensors.

As mentioned earlier, the goal of this work is to prove the feasibility of combining visual information with deep network to estimate the applied forces during robotic-assisted surgeries. However, when we talk about haptics technology in RAMIS settings, we have to consider two questions. One is how to acquire the significant information, which we talked here in this paper offering an efficient force estimation approach, and the other is how to transmit that information to the surgeon. In other words, how to display the estimated forces to the surgeon? While the first question was tackled in this paper, the second one is part of our future work in which we will use virtual reality to provide an efficient way to display the estimated force to the surgeon. Once the feedback to the surgeon is provided, then we will carry out a clinical oriented study and compare it against other alternatives.

ACKNOWLEDGMENTS

This work was supported by a FPU national scholarship from the Spanish Ministry of Education with reference AP2012-1943. The work has been developed in the frame of Project RTC-2015-3926-1, from MINECO and with Feder funds.

REFERENCES

- [1] E. Wilson, H. Bagshahi, and V. Woodruff, "Overview of general advantages, limitations, and strategies," *Book Chapter, Robotics in General Surgery*, Springer, 2014.
- [2] M. Shahinpoor and S. Gheshmi, "Introduction to surgical robots' general configurations," *Book Chapter, Robotic Surgery - Smart Materials, Robotic Structures, and Artificial Muscles*, Taylor & Francis Group, 2015.
- [3] M. Kroh and S. Chalikonda, "Essentials of robotic surgery," Springer, 2015.
- [4] G. Spinoglio, A. Marano, and F. G., "Robotic surgery: Current applications and new trends," Springer, 2015.
- [5] M. Diana and J. Marescaux, "Robotic surgery," *British Journal of Surgery*, pp. 15–28, 2015.
- [6] C. Pacchierotti, A. Tirmizi, and D. Prattichizzo, "Improving transparency in teleoperation by means of cutaneous tactile force feedback," *ACM Transactions on Applied Perception*, 2014.
- [7] O. Van der Meijden and M. Schijven, "The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review," *Surgical Endoscopy*, pp. 1180–1190, 2009.
- [8] C. Pacchierotti, L. Meli, F. Chinello, M. Malvezzi, and D. Prattichizzo, "Cutaneous haptic feedback to ensure the stability of robotic teleoperation systems," *The International Journal of Robotics Research*, vol. 34, no. 14, pp. 1773–1787, 2015.
- [9] T. Lendvay, B. Hannafor, and R. Satava, "Future of robotic surgery," *The Cancer Journal* Vol. 19, No. 2, pp. 109–119, 2013.
- [10] B. Bayle, M. Joinie-Maurin, L. Barbe, J. Gangloff, and M. de Mathelin, "Robot interaction control in medicine and surgery: Original results and open problems," *Book Chapter in Computational Surgery and Dual Training*, pp. 196–191, 2014.
- [11] M. Yip, S. Yuen, and R. Howe, "A robust uniaxial force sensor for minimally invasive surgery," *IEEE Transactions on Biomedical Engineering*, Vol. 57, No. 5, pp. 1008–1011, 2010.
- [12] P. Puangmali, H. Liu, K. Althoefer, and L. Seneviratne, "Miniature 3-axis distal force sensor for minimally invasive surgical palpation," *IEEE Transactions on Mechatronics*, pp. 646–656, 2012.
- [13] A. Faragasso, J. Bimbo, Y. Noh, A. Jiang, S. Sareh, H. Liu, T. Nanayakkara, H. Wurdemann, and K. Althoefer, "Novel uniaxial force sensor based on visual information for minimally invasive surgery," *IEEE International Conference on Robotics and Automation*, pp. 2934–2939, 2014.
- [14] A. Okamura, L. Verner, C. Reiley, and M. Mahvash, "Haptic feed-back in robot-assisted minimally invasive surgery," *Book Chapter in Robotics Research, Springer Tracts in Advanced Robotics*, pp. 361–372, 2011.
- [15] K. Helton, B. Ratner, and N. Wisniewski, "Biomechanics of the sensor-tissue interface-effects of motion, pressure, and design on sensor performance and the foreign body response-part I: Theoretical framework," *Journal of Diabetes Science and Technology* Vol. 5, pp. 632–646, 2011.
- [16] S. Sokhanvar, J. Dargahi, S. Najarian, and S. Arbatani, "Clinical and regulatory challenges for medical devices tactile sensing and displays," *Haptic Feedback for Minimally Invasive Surgery and Robotics*, Wiley Publications, 2012.
- [17] M. Massimino and T. Sheridan, "Sensory substitution for force feedback in teleoperation," *Analysis, Design and Evaluation of Man-Machine Systems*, 1992.
- [18] J. K. Koehn and K. J. Kuchenbecker, "Surgeons and non-surgeons prefer haptic feedback of instrument vibrations during robotic surgery," *Surgical Endoscopy*, pp. 2970–2983, 2014.
- [19] W. McMahan, J. Gewirtz, D. Standish, P. Martin, J. Kunkel, M. Lilavois, A. Wedmid, D. Lee, and K. Kuchenbecker, "Tool contact acceleration feedback for telerobotic surgery," *IEEE Transactions on Haptics*, vol. 4, no. 3, pp. 210–220, 2011.
- [20] Y. Visell and S. Okamoto, "Vibrotactile sensation and softness perception," *Book Chapter in Multisensory Softness*, Springer, pp. 31–47, 2014.
- [21] C. Pacchierotti, D. Prattichizzo, and K. Kuchenbecker, "Cutaneous feedback of fingertip deformation and vibration for palpation in robotic surgery," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 2, pp. 278–287, 2016.
- [22] Z. F. Quek, S. Schorr, I. Nisky, W. Provancher, and A. Okamura, "Sensory substitution and augmentation using 3-degree-of-freedom skin deformation feedback," *IEEE Transactions Haptics*, vol. 8 num. 2, pp. 209–221, 2015.
- [23] B. Bethea, K. M. Okamura, A.M., T. Fitton, S. Cattaneo, V. Gott, W. Baumgartner, and Y. D.D., "Application of haptic feedback to robotic surgery," *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 2004.
- [24] M. Greminger and B. Nelson, "Vision-based force measurement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pp. 290–298, 2004.
- [25] J. Kim, F. Janabi-Sharifi, and J. Kim, "A haptic interaction method using visual information and physically based modeling," *IEEE Transactions on Mechatronics*, Vol. 15, No. 4, pp. 636–645, 2010.
- [26] E. Noohi, S. Parastegari, and M. Zefran, "Using monocular images to estimate interaction forces during minimally invasive surgery," *IEEE International Conference on Intelligent Robots and Systems*, pp. 4297–4302, 2014.
- [27] W. Kim, S. Seung, K. Choi, S. Park, S. Ko, and J. Park, "Image-based force estimation of deformable tissue using depth map for single-port surgical robot," *International Conference on Control, Automation and Systems*, pp. 1716–1719, 2012.
- [28] M. Ammi, H. Ladjal, and A. Ferreira, "Evaluation of 3d pseudo-haptic rendering using vision for cell micromanipulation," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.
- [29] A. Kerdok, S. Cotin, M. Ottensmeyer, A. Galea, R. Howe, and S. Dawson, "Truth cube: Establishing physical standards for soft tissue simulation," *Medical Image Analysis*, 2003.
- [30] M. Greminger and B. Nelson, "Modeling elastic objects with neural networks for vision-based force measurement," *IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, pp. 1278–1283, 2003.
- [31] F. Karimirad, S. Chauhan, and B. Shirinzadeh, "Vision-based force measurement using neural networks for biological cell microinjection," *Journal of Biomechanics*, Volume 47, pp. 1157–1163, 2014.
- [32] A. Mozaffari, S. Behzadipour, and M. Kohani, "Identifying the tool-tissue force in robotic laparoscopic surgery using neuro-evolutionary fuzzy systems and a synchronous self-learning hyper level supervisor," *Elsevier Journal in Applied Soft Computing*, pp. 1278–1283, 2014.
- [33] A. Aviles, A. Marban, P. Sobrevilla, J. Fernandez, and A. Casals, "A recurrent neural network approach for 3d vision-based force estimation," *IEEE International Conference on Image Processing Theory, Tools and Applications*, 2014.
- [34] A. Aviles, S. M. Alsaleh, P. Sobrevilla, and A. Casals, "Sensorless force estimation using a neuro-vision-based approach for robotic-

- assisted surgery," *IEEE EMBS International Conference on Neural Engineering*, pp. 86–89, 2015.
- [35] A. Aviles, S. Alsaleh, P. Sobrevilla, and A. Casals, "Force-feedback sensory substitution using supervised recurrent learning for robotic-assisted surgery," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
 - [36] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.
 - [37] A. Bjorck, "Numerical methods for least squares problems," *Society for Industrial and Applied Mathematics*, 1996.
 - [38] J. Hadamard, "Lectures on the cauchy problems in linear partial differential equations," *Yale University Press, New Haven*, 1923.
 - [39] S. M. Alsaleh, A. I. Aviles, P. Sobrevilla, A. Casals, and J. Hahn, "Automatic and robust single-camera specular highlight removal in cardiac images," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
 - [40] P. Huber, "Robust regression: Asymptotics, conjectures and monte carlo," *Annals of Statistics*, 1973.
 - [41] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, 1944.
 - [42] D. Marquardt, "An algorithm for least-squares estimation of non-linear parameters," *SIAM Journal on Applied Mathematics*, pp. 431–441, 1963.
 - [43] M. Hermans and B. Schrauwen, "Training and analysing deep recurrent neural networks," *Advances in Neural Information Processing Systems* 26, 2013.
 - [44] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, 2013.
 - [45] P. Hunter, A. McCulloch, and H. ter Keurs, "Modelling the mechanical properties of cardiac muscle," *Progress in Biophysics and Molecular Biology*, vol. 69, no. 23, pp. 289 – 331, 1998.
 - [46] Y.-C. Fung, "Mathematical representation of the mechanical properties of the heart muscle," *Journal of Biomechanics*, vol. 3, no. 4, pp. 381 – 404, 1970.
 - [47] P. Mountney, D. Stoyanov, and G.-Z. Yang, "Three-dimensional tissue deformation recovery and tracking: Introducing techniques based on laparoscopic or endoscopic images," *IEEE Signal Processing Magazine*, pp. 14–24, 2010.
 - [48] E. Roche, R. Wohlfarth, J. Overvelde, N. Vasilyev, F. Pigula, D. Mooney, K. Bertoldi, and C. Walsh, "A bioinspired soft actuated material," *Advanced Materials*, pp. 1–7, 2013.
 - [49] S. McKinley, A. Garg, S. Sen, R. Kapadia, A. Murali, K. Nichols, S. Lim, S. Patil, P. Abbeel, A. Okamura, and K. Goldberg, "A single-use haptic palpation probe for locating subcutaneous blood vessels in robot-assisted minimally invasive surgery," *IEEE International Conference on Automation Science and Engineering CASE*, 2015.
 - [50] C. Pacchierotti, *Cutaneous Haptic Feedback in Robotic Teleoperation*, ser. 2192-2977. Springer International Publishing, 2015.
 - [51] M. Li, H. Liu, J. Li, L. Seneviratne, and K. Althoefer, "Tissue stiffness simulation and abnormality localization using pseudo-haptic feedback," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 5359–5364.
 - [52] S. DiMaio and S. Salcudean, "Needle insertion modeling and simulation," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 5, pp. 864–875, 2003.
 - [53] C. Wagner, N. Stylopoulos, and R. Howe, "The role of force feedback in surgery: analysis of blunt dissection," in *Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2002. HAPTICS 2002. Proceedings. 10th Symposium on, 2002, pp. 68–74.
 - [54] M. Kitagawa, D. Dokko, A. M. Okamura, and D. D. Yuh, "Effect of sensory substitution on suture-manipulation forces for robotic surgical systems," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 129, no. 1, pp. 151 – 158, 2005.